

**Всероссийская конференция
"Нелинейная динамика в когнитивных исследованиях"
Нижний Новгород, НОЦ ИПФ РАН, 13-15 мая 2009 г.**

ПРОСТАЯ НЕЙРОСЕТЕВАЯ МОДЕЛЬ УЗНАВАНИЯ

А.Т. Терехин, Е.В. Будилова, Л.М. Качалова

**Москва
Биологический факультет МГУ
Институт когнитивной нейрологии СГА**

АНАТОЛИЙ ТИМОФЕЕВИЧ ТЕРЕХИН



Профессор
кафедры общей экологии
Биологического факультета
МГУ им. М.В. Ломоносова
доктор биологических наук

Адрес: 119992, Россия, Москва
Ленинские горы 1, строение 12
Биологический факультет МГУ
кафедра общей экологии

Телефон: +7(495)939-53-64

E-mail: terekhin_a@mail.ru

<http://ecology.genebee.msu.ru>

Блокирование памяти

(tip-of-the-tongue state – «вертится на языке»)

Чехов А.И. Лошадиная фамилия. *Петербургская газета, 1885.*

James W. The principles of Psychology. *New York: Holt, 1890.*

Характерные особенности блокирования:

- 1) Парадоксальный контраст между уверенностью в том, что образ знаком, и невозможностью воспроизвести его полностью.
- 2) Повышение риска блокирования с увеличением возраста.
- 3) Более частое блокирование имен собственных.

Узнавание и воспроизведение

Нейросетевое моделирование процесса узнавания основано на том факте, что функция энергии сети Хопфилда¹ принимает большие значения для уже запомненных образов по сравнению с вновь предъявляемыми².

¹ *Hopfield J.J.* Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, 1982, v. 79(8), 2554–2558.

² *Amit D.J.* Modeling brain function—the world of attractor neural networks. *Cambridge: Cambridge University Press*, 1989.

Динамика состояний сети Хопфилда $x_i(t)$ определяется правилом Мак-Каллока и Питтса¹

$$x_i(t+1) = \text{sign} \left(\sum_{j=1}^N w_{ij}(t) x_j(t) \right)$$

а динамика синаптических весов $w_{ij}(t)$ — правилом Хебба²

$$w_{ij}(t+1) = w_{ij}(t) + x_i(t)x_j(t)$$

¹ ***McCulloch W. S., Pitts W. A logical calculus of the ideas immanent in neurons activity. Bull. Math. Biophys., 1943, v. 5, 115–133.***

² ***Hebb D.O. The Organization of Behavior. A Neuropsychological Theory. New York: Wiley, 1949.***

Функция энергии сети Хопфилда задается выражением:

$$E(X_t) = -0.5 \sum_{i=1}^N \sum_{j=1}^N w_{ij}(t) x_i(t) x_j(t)$$

Максимальное число образов, которые могут быть предъявлены сети Хопфилда и затем с ошибкой менее 1% узнаны как знакомые равно $0.023N^2$ против числа $0.145N$, которые могут быть запомнены сетью с возможностью последующего полного воспроизведения¹.

¹ *Bogacz R., Brown M.W., Giraud-Carrier C. Model of Familiarity Discrimination in the Perirhinal Cortex. J. Comput. Neurosc., 2001, v. 10(1), 5-23.*

Мы модифицировали формулу энергии, заменив внутреннюю сумму ее знаком:

$$E^*(X_t) = -0.5 \sum_{i=1}^N x_i(t) \operatorname{sign} \left(\sum_{j=1}^N w_{ij}(t) x_j(t) \right)$$

Учитывая правило Мак-Каллока и Питтса

$$x_i(t+1) = \operatorname{sign} \left(\sum_{j=1}^N w_{ij}(t) x_j(t) \right)$$

получаем:

$$E^*(X_t) = -0.5 \sum_{i=1}^N x_i(t) x_i(t+1)$$

т.е. модифицированная энергия просто пропорциональна скалярному произведению двух последовательных состояний сети.

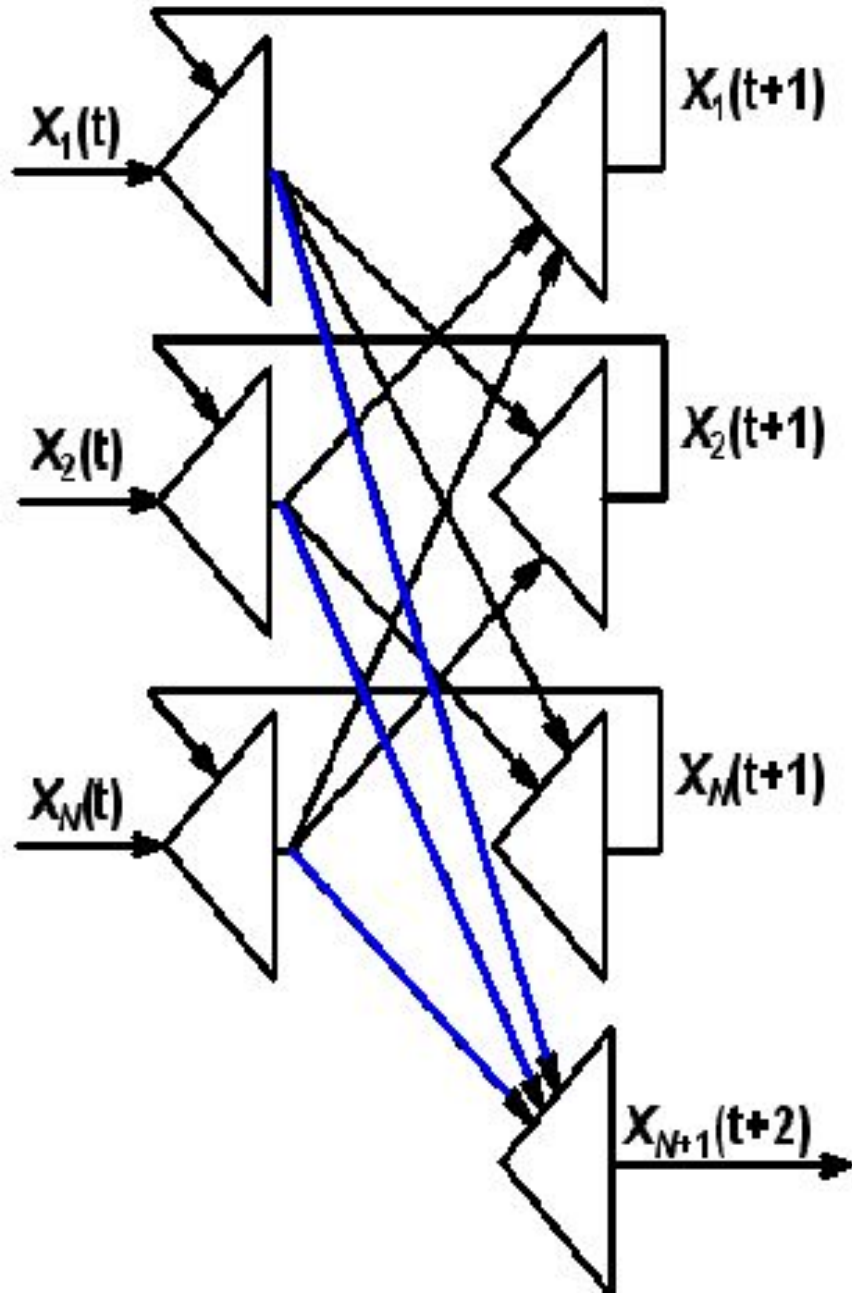
Учитывая простоту формулы

$$E^*(X_t) = -0.5 \sum_{i=1}^N x_i(t)x_i(t+1)$$

и быстроту вычислений по ней (один временной шаг) можно, во-первых, предположить, что такого рода механизм узнавания реализуется в реальных биологических нейронных сетях, и, во-вторых, предложить этот механизм для реализации функции узнавания в искусственных когнитивных системах.

Была построена конкретная нейронная сеть, решающая задачу узнавания с помощью вычисления модифицированной энергии и была определена емкость ее памяти.

Архитектура сети узнавания



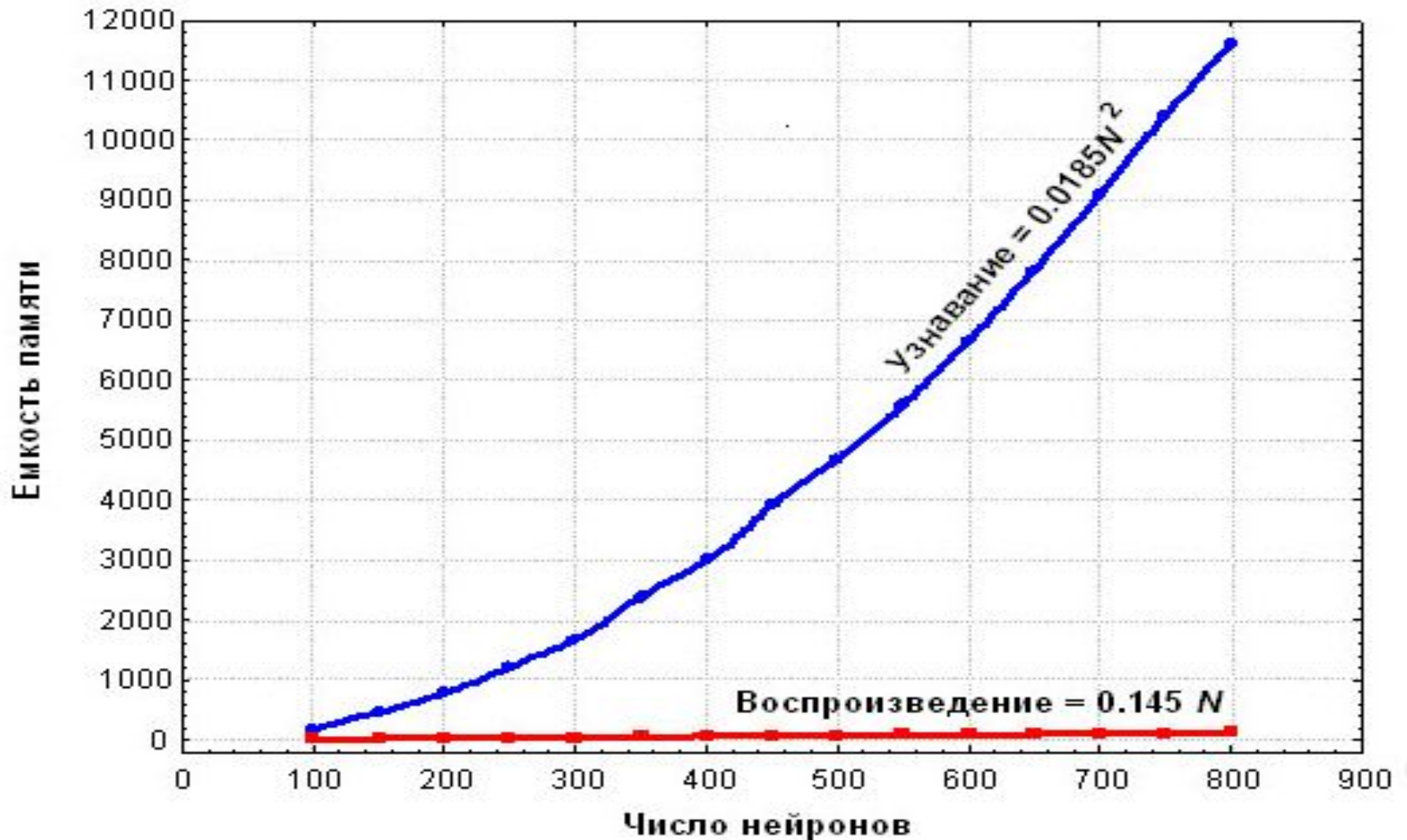
На шаге t веса от нейронов первого слоя сети Хопфилда к узнающему нейрону устанавливаются (по правилу Хебба) равными $x_1(t)$, $x_2(t)$, ..., $x_N(t)$, а на шаге $t+1$ – равными (по правилу Мак-Каллоха и Питтса) - знаку скалярного произведения векторов $x_1(t)$, $x_2(t)$, ..., $x_N(t)$ и $x_1(t+1)$, $x_2(t+1)$, ..., $x_N(t+1)$.

Вычисление емкости памяти узнавания

```
N=700; i=0;
for P=500:500:12000
i=i+1;
Xfam=sign(rand(N,P)-.5);
Xnov=sign(rand(N,P)-.5);
W=Xfam*Xfam'-P*eye(N);
Yfam=sign(W*Xfam);
Ynov=sign(W*Xnov);
for p=1:P
Efam(p)=Xfam(:,p)'*Yfam(:,p);
Enov(p)=Xnov(:,p)'*Ynov(:,p);
end;
PP(i)=P;
Mf(i)=mean(Efam);
Sf(i)=std(Efam);
Mn(i)=mean(Enov);
Sn(i)=std(Enov);
end;
plot(PP,Mf,PP,Mf-2.33*Sf,PP,Mn,PP,Mn+2.33*Sn);
```

Емкость памяти узнавания

Емкость памяти узнавания построенной сети равна $0.018N^2$, т.е. 80% от максимума $0.023N^2$.



Блокирование памяти

(tip-of-the-tongue state – «вертится на языке»)

Чехов А.И. Лошадиная фамилия. *Петербургская газета, 1885.*

James W. The principles of Psychology. *New York: Holt, 1890.*

Характерные особенности блокирования:

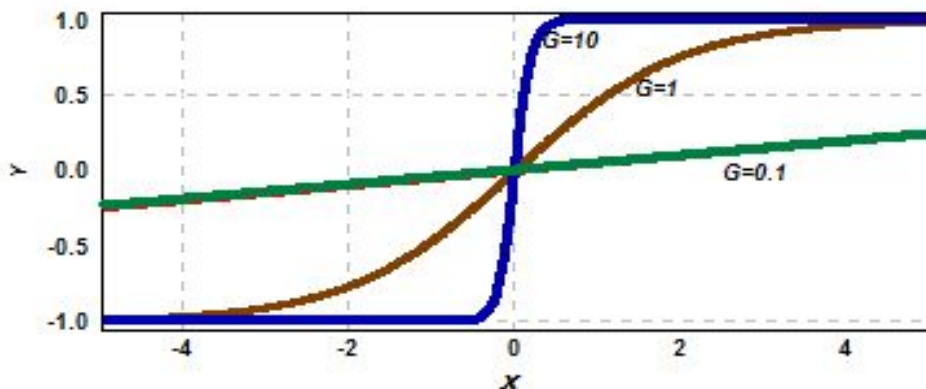
- 1) Парадоксальный контраст между уверенностью в том, что образ знаком, и невозможностью воспроизвести его полностью.
- 2) Повышение риска блокирования с увеличением возраста.
- 3) Более частое блокирование имен собственных.

Модель старения мозга

С возрастом в мозге происходит множество нейроанатомических и нейрохимических изменений, способствующих ослаблению межнейронных связей. Например, начиная с 20-летнего возраста постоянно снижается плотность многих постсинаптических рецепторов, вследствие чего снижается чувствительность нейронов к входящим сигналам.

Исходя из этого, мы ввели в сигмоидную функцию активации параметр G и связали процесс старения мозга с уменьшением величины этого параметра.

$$x_i = \frac{2}{1 + e^{-G \sum w_{ij} x_j}} - 1$$



**Функция энергии для сети Хопфилда
является суммой двух членов:**

$$E(x_1, x_2, \dots, x_n) = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n x_i w_{ij} x_j + \frac{1}{G} \sum_{i=1}^n \ln[(1+x_i)^{1+x_i} (1-x_i)^{1-x_i}]$$

**При больших G преобладающее значение имеет
первый член:**

$$E_1 = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n x_i w_{ij} x_j$$

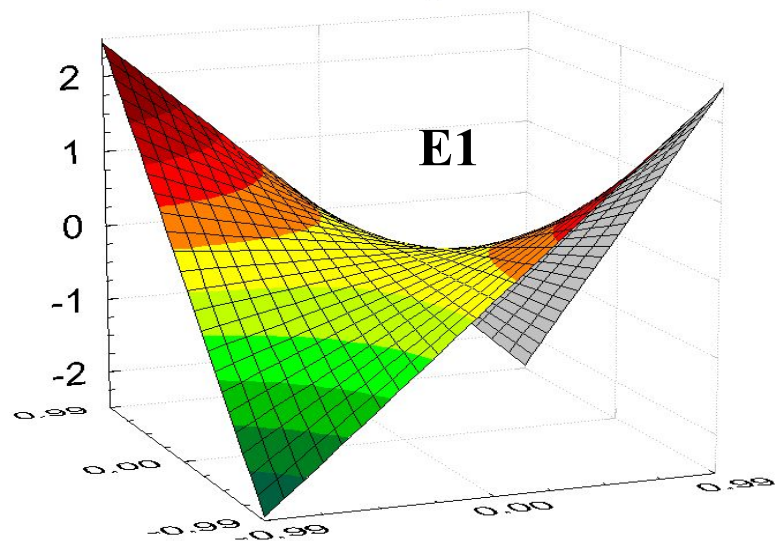
При малых G преобладает второй член:

$$E_2 = \frac{1}{G} \sum_{i=1}^n \ln[(1+x_i)^{1+x_i} (1-x_i)^{1-x_i}]$$

Функция энергии для сети из двух нейронов

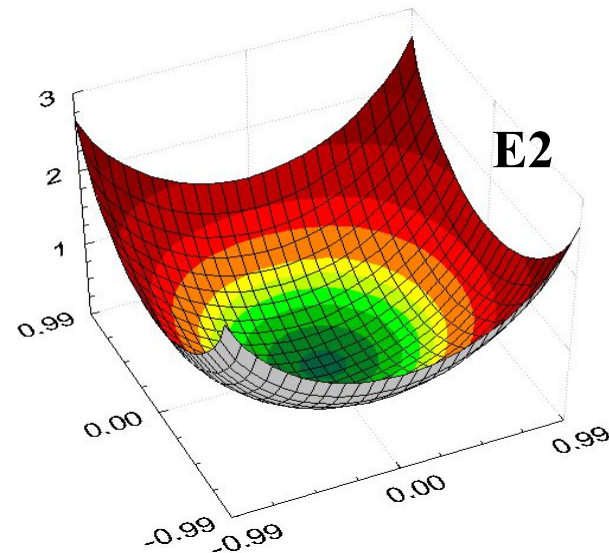
$$z = -2.5 * x * y$$

E1



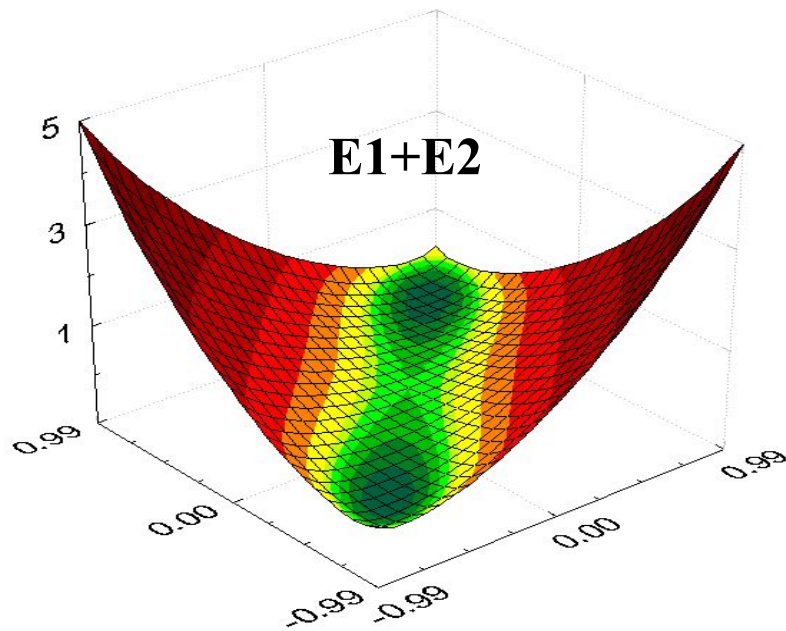
$$z = \log((1+x)^{(1+x)}) + \log((1-x)^{(1-x)}) + \log((1+y)^{(1+y)}) + \log((1-y)^{(1-y)})$$

E2

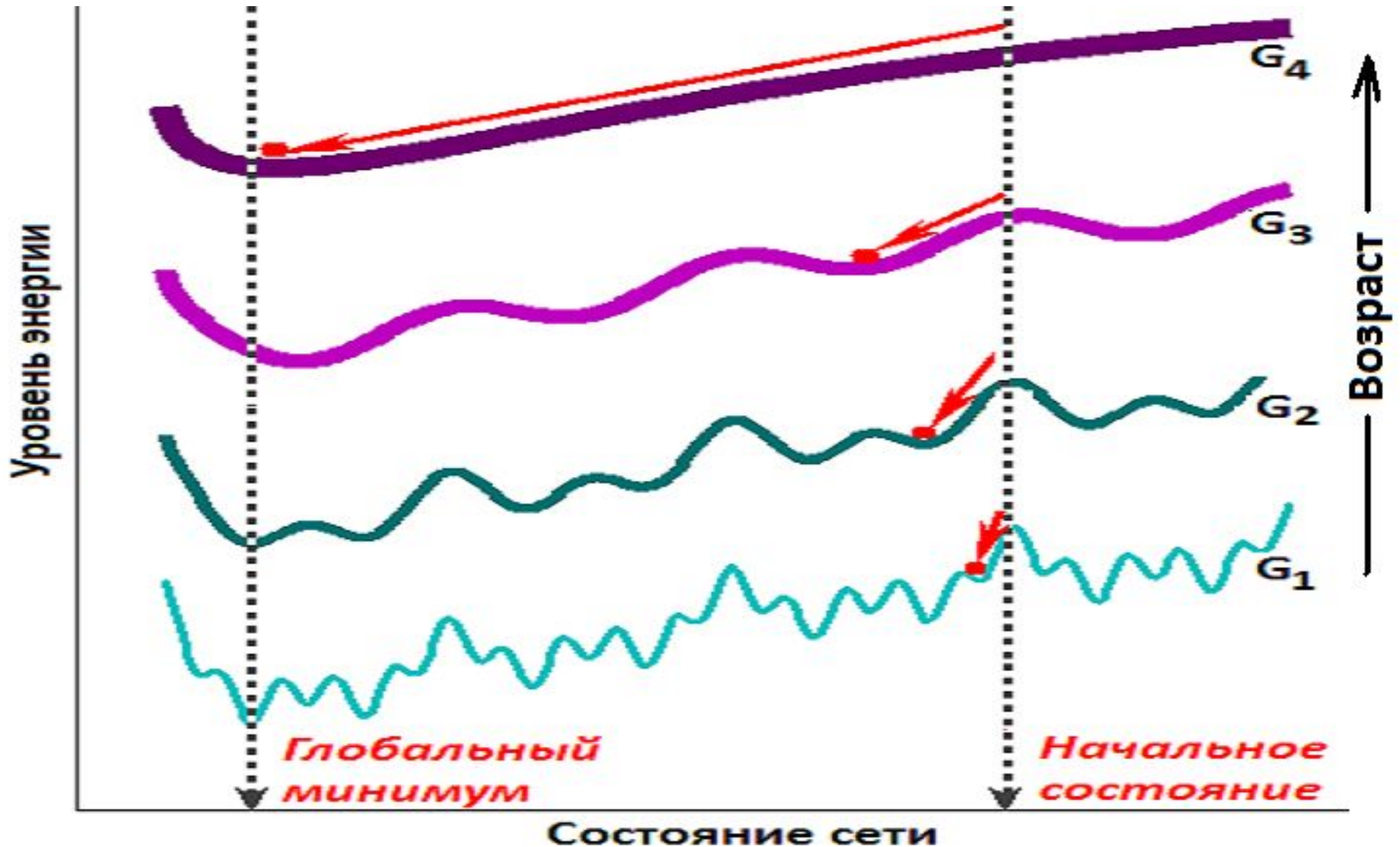


$$z = -2.5 * x * y + \log((1+x)^{(1+x)}) + \log((1-x)^{(1-x)}) + \log((1+y)^{(1+y)}) + \log((1-y)^{(1-y)})$$

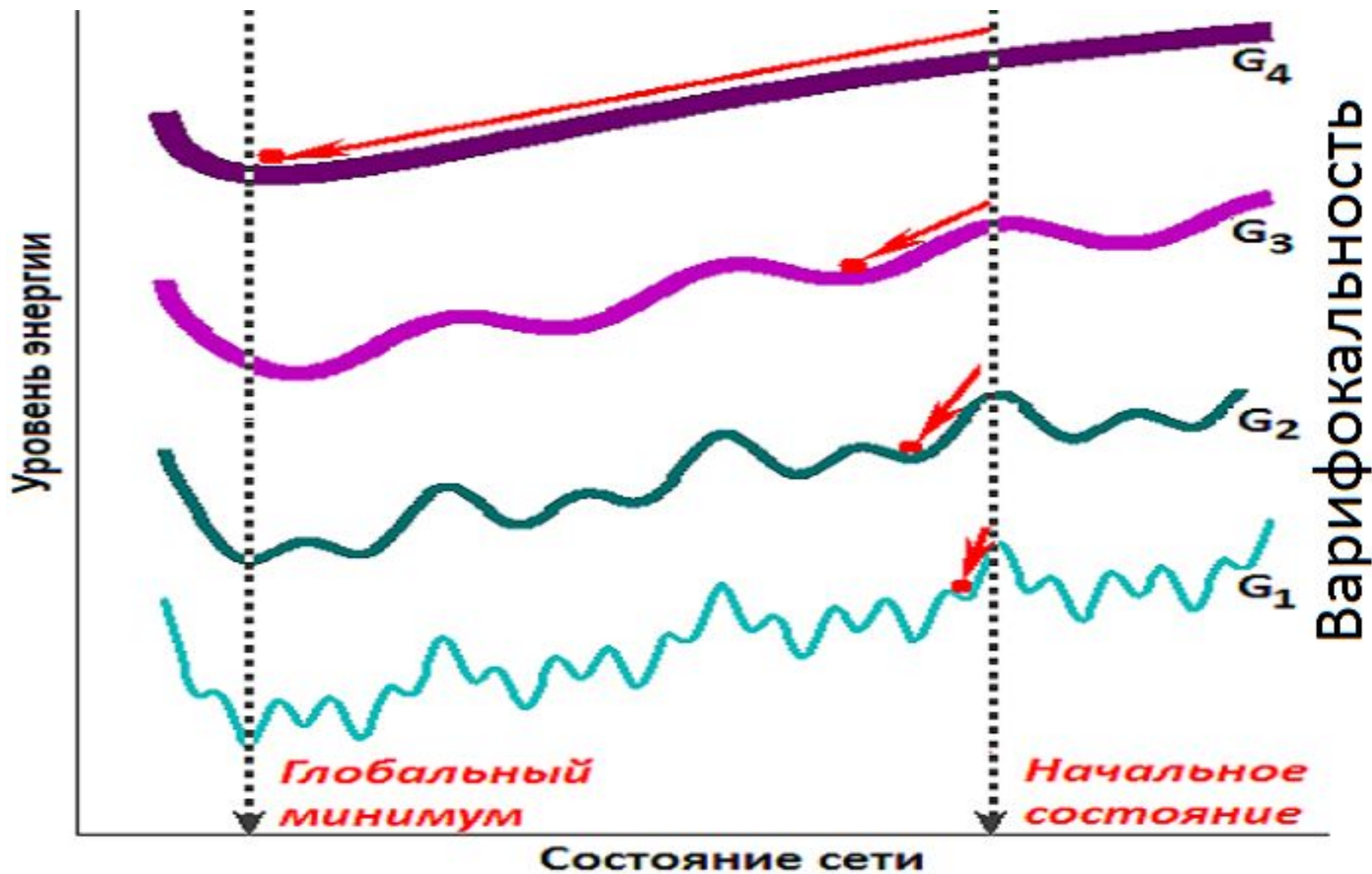
E1+E2



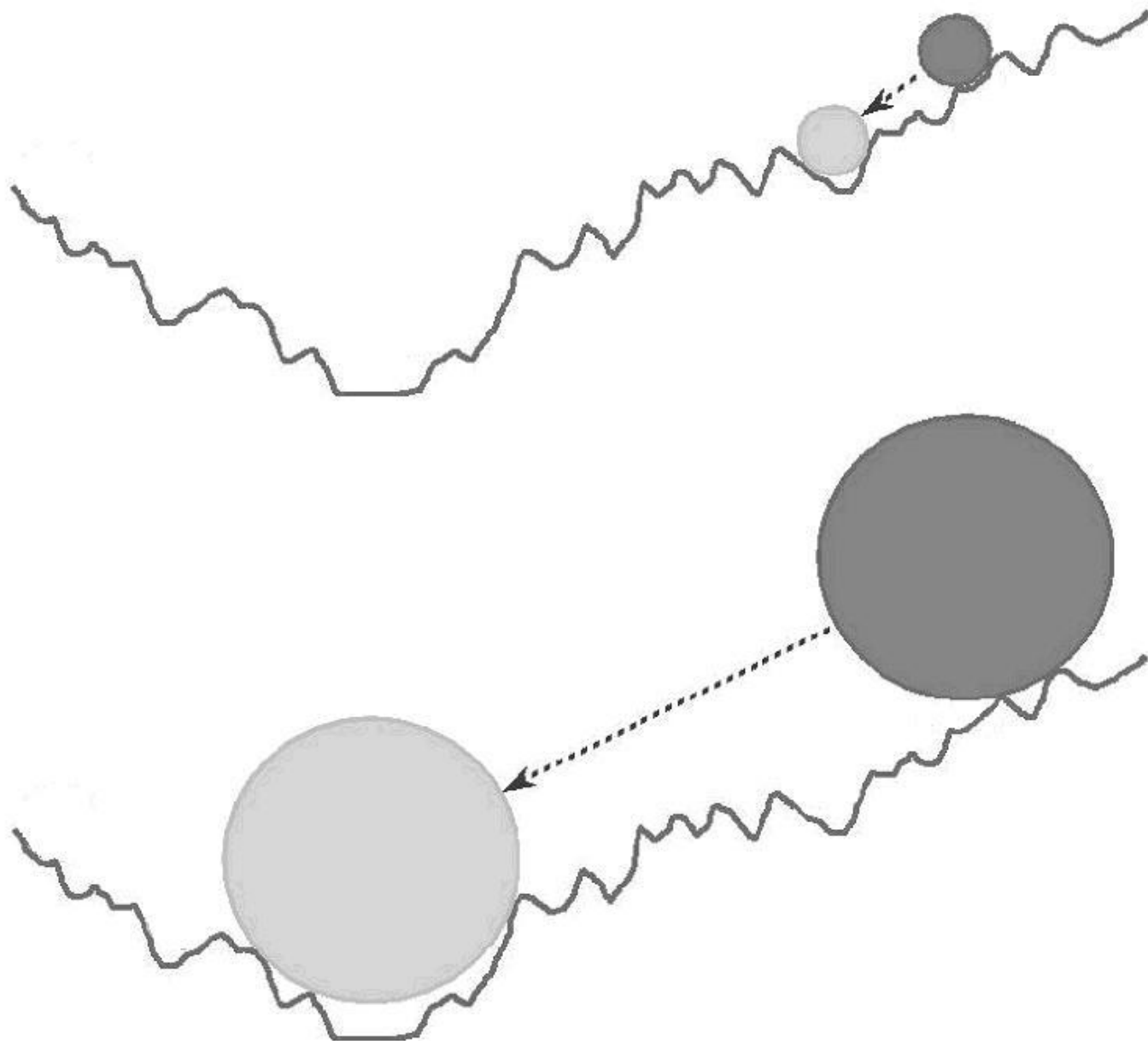
Когнитивный эффект возрастного сглаживания функции энергии сети ($G_1 > G_2 > G_3 > G_4$)



«Варифокальность» мышления



«Варифокальность» мышления



Карпенко М.П., Качалова Л.М., Будилова Е.В., Терехин А.Т. Когнитивные преимущества третьего возраста: нейросетевая модель старения мозга. Журнал высшей нервной деятельности, 2009, т. 59(2), 291-295.

Карпенко М.П., Чмыхова Е.В., Терехин А.Т. Модель возрастного изменения восприятия времени. Вопросы психологии, 2009, 2, 75-81.

Будилова Е.В., Карпенко М.П., Качалова Л.М., Терехин А.Т. Узнавание и воспроизведение: нейросетевая модель. Биофизика, 2009, т. 54, № 3.

terekhin_a@mail.ru

<http://ecology.genebee.msu.ru>